AFHRL-TP-87-61

DTIC FILE COPY

(2)

AIR FORCE

HUMAN RESOURCES

AD-A194 753

VALIDITY OF THE ACADEMIC APTITUDE COMPOSITE OF THE AIR FORCE OFFICER QUALIFYING TEST (AFOQT)

Darrell D. Hartke
Lawrence O. Short, Lt Col, USAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

April 1988
Final Technical Paper for Period October 1985 - December 1987

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

88 5 16 098

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-87-61	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b. OFFICE SYMBOL (If applicable) AFHRL/MO	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62703F; PROJECT NO. 7719; TASK NO 18; WORK UNIT ACCESSION NO. 19

11. TITLE (Include Security Classification)

Validity of the Academic Aptitude Composite of the Air Force Officer Qualifying Test (AFOQT)

12. PERSONAL AUTHOR(S)

Hartke, D.D.; Short, L.O.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM Oct 85 TO Dec 87	April 1988	14

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

academic aptitude composite; Air Force Officer Qualifying Test (AFOQT); meta-analysis; validity generalization

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This paper addresses the use of the Schmidt-Hunter meta-analysis procedure to determine the generalizability of the Air Force Officer Qualifying Test (AFOQT) validities across Air Force specialties (AFSs). Meta-analyses were conducted on all available Academic Aptitude composite validities aggregated and disaggregated into major occupational subgroupings. The results suggest that although the validity of the Academic Aptitude composite of the AFOQT may vary across AFSs, the AFOQT is of general value in officer selection.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b. TELEPHONE (Include Area Code) (512) 536-3877 22c. OFFICE SYMBOL AFHRL/TSR

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

AFHRL Technical Paper 87-61   April 1988

VALIDITY OF THE ACADEMIC APTITUDE COMPOSITE OF THE AIR FORCE OFFICER QUALIFYING TEST (AFOQT)

Darrell D. Hartke
Lawrence O. Short, Lt Col, USAF

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

Lonnie D. Valentine
Chief, Force Acquisition Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

Schmidt-Hunter meta-analyses were conducted on 47 Academic Aptitude composite validities to determine the degree to which validity generalized across non-rated AFSs. Analyses were conducted on a full set of 47 validity coefficients and on validity coefficients within 4 occupational subgroups. In only one subgroup, intelligence and security police specialties, was validity generalization supported. However, the results did support the usefulness of the Academic Aptitude composite for Air Force officer selection.

PREFACE

This work was completed under Task 771918, Selection and Classification Technologies, which is part of a larger effort in Force Acquisition and Distribution. It was subsumed under Work Unit 77191819, "Development and Validation of Selection Methodologies." This work unit was established in response to Air Force Regulation (AFR) 35-8.

TABLE OF CONTENTS

LIST OF TABLES

VALIDITY OF THE ACADEMIC APTITUDE COMPOSITE OF THE
AIR FORCE OFFICER QUALIFYING TEST (AFOQT)

I. INTRODUCTION

The Air Force Officer Qualifying Test (AFOQT) is a battery of tests measuring aptitude for selection into officer commissioning and post-commissioning training programs. Test versions of the AFOQT, in use since 1981, consist of 380 items organized into 16 subtests which combine to produce 5 composites: Verbal (V), Quantitative (Q), Academic Aptitude (AA), Pilot (P), and Navigator-Technical (N-T) (see Table 1). The V and Q composites are used for selection purposes, and AA is simply an additive combination of these two. P and N-T are further used to select candidates for undergraduate pilot and navigator training, respectively.

Table 1. Subtests in AFOQT Aptitude Composites

	Composite				
Subtest	Verbal	Quantitative	Academic Aptitude	Pilot	Navigator-Technical
Verbal Analogies	X		X	X	
Arithmetic Reasoning		X	X		X
Reading Comprehension	X		X		
Data Interpretation		X	X		X
Word Knowledge	X		X		
Math Knowledge		X	X		X
Mechanical Comprehension				X	X
Electrical Maze				X	X
Scale Reading				X	X
Instrument Comprehension				X	
Block Counting				X	X
Table Reading				X	X
Aviation Information				X	
Rotated Blocks					X
General Science					X
Hidden Figures					X

Note. Table describes the composition of AFOQT forms in use since 1981.

The AFOQT first became operational in 1953, and has undergone several revisions since that time. With each of these revisions, the nature of the AFOQT has changed. Some subtests have been deleted from the battery, others have been added, and the names and composition of the composites have changed. Nevertheless, the need for 3-year revision cycles necessitates that validity be as generalizable as possible across both test forms and Air Force specialties (AFSs), especially non-rated specialties. Thus, this study had two purposes: (a) to determine if the validity of the AFOQT generalized across all forms of the instrument, in spite of the alterations that have occurred; and (b) to determine if the validity of the AFOQT generalized across various non-flying AFSs. In this context, validity is defined by the correlation between the AFOQT and grades obtained in officer technical training programs. To investigate these questions, a series of AFOQT validity studies were subjected to a Schmidt-Hunter meta-analysis procedure (Hunter, Schmidt, & Jackson, 1982).

II. THE SCHMIDT AND HUNTER META-ANALYSIS PROCEDURE

There are several meta-analysis procedures currently available. However, the procedure developed by Schmidt and Hunter has advantages which make it the best choice here. First, when

accumulating results from a number of studies, the Schmidt and Hunter procedure produces an overall effect size (usually stated as a correlation) rather than a cumulative probability. Procedures which attempt to find only a cumulative probability (see Mosteller & Bush, 1954) provide no information about the magnitude of the effect associated with the studies. As was noted by Hunter et al. (1982), "the practical and theoretical implications of an effect depend at least as much on its size as on its existence" (p. 133).

Second, unlike other meta-analysis procedures, the Schmidt and Hunter procedure does not take the observed variance in effect sizes at face value. Hunter et al. (1982) argued that the observed variance in effect sizes across studies may result from two major sources. One source of variance is moderator variables, variables which cause differences in the correlation between two other variables. These variables often produce substantive differences in results across studies. The other source of variance is artifactual. Hunter et al. (1982) identified five "artifacts" which may produce variance in the results across studies: sampling error; study differences in range restriction; study differences in the reliability of measurement; study differences in instrument validity; and computational, typographical, and transcription errors. Although the last two artifacts are as yet uncontrollable, Hunter et al. (1982) provided formulas for estimating the variance across studies due to the first three artifacts.

When accumulating effect sizes across studies using the Schmidt and Hunter procedure, one must first determine the extent to which the observed variance in effect sizes across studies is due to artifacts. This is done by calculating a weighted mean validity coefficient corrected for study differences in range restriction and predictor/criterion reliability, the observed variance in the set of effect sizes, and an estimate of the amount of variance in the set of effect sizes due to sampling error. If variance due to sampling error accounts for all (or a large part) of the observed variance in the set of effect sizes, it is assumed there is no difference in the true validities across the set of studies. The weighted mean validity coefficient is then considered the best estimate of the true validity. The search for moderator variables should be pursued only if substantial variance exists above and beyond that due to sampling error.

Based on previous research by Schmidt and Hunter (Hunter et al., 1982; Schmidt & Hunter, 1981), a 75% decision rule has been suggested as a guide for interpreting meta-analysis results. That is, if variance due to sampling error accounts for more than 75% of the observed variance in results across studies, it is assumed that the additional 25% of observed variance is due to artifacts which are not correctable (i.e., study differences in instrument validity; computational, typographical, and/or transcription errors), and that no substantive variance exists. If variance due to sampling error accounts for less than 75% of the observed variance in results across studies, the possibility of moderator variable effects exists.

III. METHOD

A thorough review of both published and unpublished, historical and current literature on the validity of the AFOQT produced three studies which reported validity coefficients of the Academic Aptitude composite in predicting technical school success (academic grades) for non-flying AFSs. The review was limited to Academic Aptitude composite validities since this composite contains all subtests used in selecting officers into non-rated AFSs. These three studies (Arth, 1986; Finegold & Rogers, 1985; Miller, 1960) produced a total of 47 independent validity coefficients covering a number of AFSs.

The manner in which results were reported in these three studies severely limited the conduct of this meta-analysis. First, in the Arth (1986) study, predictor information (AFOQT scores) was collected using several different forms of the AFOQT (Forms L, M, N, and O). Therefore, the correlations reported do not represent the relationship between training school success and

scores on any one form of the AFOQT. In the Finegold and Rogers (1985) and the Miller (1960) studies, there is no indication of which form, or forms, of the AFOQT was used in collecting predictor information. Therefore, as in the Arth (1986) study, the correlations may not represent the relationship between training school success and any single form of the AFOQT. This being the case, it was not possible, using the data available, to determine the generalizability of AFOQT validities across various forms of the test.

Second, the three studies included in this meta-analysis failed to report data concerning predictor and criterion reliabilities, as well as the variance of AFOQT scores. This lack of information precluded correcting validity coefficients for differences in predictor/criterion reliability and differences in range restriction. Therefore, the reported validities were corrected for only one artifact: sampling error.

In light of these restrictions, the meta-analysis was conducted in the following way. First, all 47 validity coefficients were analyzed to determine if the AFOQT Academic Aptitude composite's validity generalized across all jobs represented in the sample. Next, the validity coefficients were separated into subgroups based on the first digit of the AFS to define broadly based occupational groups. Subgroups containing six or more validity coefficients were then analyzed separately to determine if Academic Aptitude composite validity generalized within the major occupational groupings. Four subgroups with six or more validity coefficients were identified: (a) air traffic controllers, air weapons controllers, and air weapons directors; (b) communications specialties; (c) resources management specialties; and (d) intelligence and security police specialties. Table 2 provides a full list of the 47 validity coefficients and their occupational subgroup placement.

For both the full group and subgroup meta-analyses, if less than 75% of the observed variance in validity coefficients across studies was attributable to sampling error, it was concluded that one or more variables other than occupational category was moderating the validity coefficients. Consequently, validity could not be generalized across AFSs.

IV. RESULTS

The results of the full set and subgroup analyses are presented in Table 3. The meta-analysis on the full set of 47 validity coefficients produced a weighted mean correlation of .39, with 33% of the observed variance attributable to sampling error. Therefore, it was concluded that the true validity was not the same across all occupations included in this analysis. Four subgroup analyses were conducted. For subgroup 1 (air traffic controllers, air weapons controllers, and air weapons directors), the weighted mean correlation was .34, with sampling error accounting for 69% of the observed variance. The weighted mean correlation for subgroup 2 (communications specialties) was .40, and sampling error accounted for 56% of the observed variance. The weighted mean correlation for subgroup 3 (resources management specialties) was .37, with 44% of the observed variance accounted for by sampling error. Finally, the meta-analysis of validities in subgroup 4 (intelligence and security police) produced a weighted mean correlation of .44, with 100% of the observed variance accounted for by sampling error.

In one case (subgroup 4) the meta-analysis results indicate that sampling error accounted for more than 75% of the observed variance around the weighted mean validity coefficient (.44). For only this one subgroup can it be concluded that the Academic Aptitude composite's validity is the same for all subgroup occupations and that observed variance in the reported validities is due to artifacts.

Table 2. Validity Coefficients Used in the Meta-Analysis Procedure

Study	Air Force Specialty Code (AFSC)[a]	r	n	Occupational subgroup
Miller 1960	Communications -	.55	84	2
Miller 1960	Aircraft Maintenance -	.52	79	*
Miller 1960	Aircraft Maintenance -	.58	164	*
Miller 1960	Air Transport -	.29	76	*
Miller 1960	Surface Transport -	.42	70	*
Miller 1960	Supply -	.38	164	3
Miller 1960	Supply -	.52	125	3
Miller 1960	Personnel -	.48	116	*
Miller 1960	Air Police -	.31	97	4
Finegold & Rogers 1985	Air Weapons Controller (17XX)	.35	986	1
Arth 1986	Air Traffic Controller (1631)	.50	91	1
Arth 1986	Air Weapons Director (1741A)	.31	217	1
Arth 1986	Air Weapons Director (1741D)	.41	109	1
Arth 1986	Air Weapons Director (1741X)	.34	593	1
Arth 1986	Air Weapons Director (1744A)	.17	120	1
Arth 1986	Missile Operations (1821F)	.55	456	*
Arth 1986	Space Systems (2001)	.43	185	*
Arth 1986	Space Systems (2031)	.36	145	*
Arth 1986	Weather (2524)	.08	78	*
Arth 1986	Communications-Electronics (3016)	.28	97	2
Arth 1986	Communications-Electronics (3021)	.44	382	2
Arth 1986	Communications-Electronics (3024D)	.47	113	2
Arth 1986	Communications-Electronics (3031)	.41	326	2
Arth 1986	Communications-Electronics (3051)	.28	215	2
Arth 1986	Aircraft Maintenance and Munitions (4021)	.31	850	*
Arth 1986	Aircraft Maintenance and Munitions (4051A)	.48	364	*
Arth 1986	Aircraft Maintenance and Munitions (4054X)	.05	98	*
Arth 1986	Computer Systems (5131B)	.49	308	*
Arth 1986	Computer Systems (5135B)	.33	89	*
Arth 1986	Transportation (6051)	.52	354	3
Arth 1986	Services (6221)	.26	186	3
Arth 1986	Supply Management (6421)	.35	324	3
Arth 1986	Supply Management (6424)	.33	103	3
Arth 1986	Acquisition Contracting/Manufacturing (6531)	.41	248	3
Arth 1986	Acquisition Contracting/Manufacturing (6534)	.17	109	3
Arth 1986	Logistic Plans & Programs (6221)	.31	129	3
Arth 1986	Financial (6721)	.30	114	3
Arth 1986	Financial (6731)	.27	121	3
Arth 1986	Management Analysis (6921)	.36	124	3
Arth 1986	Administration (7000)	.35	770	*
Arth 1986	Personnel (7321)	.42	292	*
Arth 1986	Manpower Management (7421)	.48	145	*
Arth 1986	Intelligence (8000)	.50	168	4
Arth 1986	Intelligence (8031)	.50	159	4
Arth 1986	Intelligence (8041)	.44	141	4
Arth 1986	Intelligence (8051)	.46	420	4
Arth 1986	Security Police (8121)	.39	286	4

[a]AFSCs from the Miller (1960) study are out of date. Therefore, to avoid confusion, they are not presented here.

*These validity coefficients were not subjected to a subgroup meta-analysis due to the small number of validity coefficients within their respective occupational subgroup.

Table 3. Results of the Full Set and Occupational Subgroup Meta-Analyses

	Number of coefficients in analysis	Weighted mean r	Observed variance in coefficients	Expected variance due to sampling error	% of Observed variance due to sampling error
Full Set	47	.39	.0093	.0031	33%
Subgroup 1	6	.34	.0032	.0022	69%
Subgroup 2	6	.40	.0062	.0035	56%
Subgroup 3	12	.37	.0097	.0042	44%
Subgroup 4[a]	6	.44	.0029	.0031	100%

[a]In this case, the variance expected due to sampling error alone was larger than the total observed variance among coefficients. This means the total observed variance is actually less than that expected due to sampling error alone and confirms that all observed variance (100%) is due to sampling error.

V. DISCUSSION AND CONCLUSIONS

The results of the meta-analysis of the full set of validity coefficients indicate that the validity of the Academic Aptitude composite is not the same across all non-rated AFSs. This is to be expected considering the wide variety of AFSs included in the analysis. Further, validity generalizability was indicated in only one of the four occupational subgroups (subgroup 4). While these results were more unexpected, three issues must be considered before accepting the existence of moderator variables which influence the validity of the Academic Aptitude composite within occupational subgroups. First, only one of the statistical artifacts, sampling error, which could contribute to variance in observed effects sizes was controlled for in this study. If other artifacts (i.e., range restriction, predictor/criterion reliability) could have been controlled for, it is possible that no substantive variance would have remained.

Second, the number of validity coefficients available for subgroup analyses was relatively small (12 coefficients for subgroup 3, and 6 for each of the other 3 subgroups). It is possible that the small number of validity coefficients in each subgroup may not have been representative of the larger population of validity coefficients. As greater numbers of validity coefficients within each occupational subgroup become available, confidence in the results of a meta-analysis would increase.

Finally, the occupational subgroups used in this study were quite broad. Subgroupings were based on the first digit of the AFS. More narrowly defined subgroups (i.e., subgroupings based on similarity in job content, task dimensions, etc.) may have more fully supported validity generalization.

Although validity generalization was not established for three of the four subgroups, this study does support the usefulness of the AFOQT in general, and the Academic Aptitude composite in particular, for Air Force officer selection. The meta-analysis on the full set of 47 validity coefficients suggests that, while the true validity of the Academic Aptitude composite varies across AFSs, these individual validities should be acceptable in most cases. The mean weighted validity coefficient, .39, is the best estimate of the average validity across all AFSs. The true validity for the individual AFSs will vary around this mean. After subtracting the variance in validities due to sampling error from the observed variance, the residual standard deviation is equal to .08. Assuming a normal distribution of validity coefficients, the value above which 90% of all the true validities will lie is approximately 1.29 standard deviations below the

weighted mean validity coefficient (see Schmidt & Hunter, 1977). Therefore, 90% of the true validities for all AFSs should be above .29.

This study was merely a first step in assessing the generalizability of the AFOQT validity, particularly the Academic Aptitude composite validity. Currently, validity research is being performed on more recent forms of the AFOQT (Forms O and P). Data from this research could be combined with the data from the three studies used in the present investigation to provide a much larger and possibly more representative data set upon which meta-analysis could be conducted. Confidence in the results of the meta-analysis would increase when using a larger data set.

Additional research into the generalizability of AFOQT validity could concentrate on a number of areas. First, efforts to define occupational subgroups could focus on task characteristics of the occupations rather than on the first digit of the AFS. These efforts would likely result in more homogeneous subgroups than those used in the present research. More homogeneous subgroups could result in more support for validity generalization across similar AFSs. Secondly, if available, the data from which this study's 47 validity coefficients were generated could be re-examined to obtain information necessary to control for range restriction and predictor/criterion reliabilities. This information would reduce the amount of variance attributable to these artifacts, thus increasing the chances for support of validity generalization. Finally, the present study concentrated solely on the Academic Aptitude composite of the AFOQT. In future research, validity data on the other composites could be subjected to meta-analysis to determine the generalizability of their validities.

If future research did not support AFOQT composite validity generalization, the question of variables which might moderate the relationship between AFOQT scores and performance should be addressed. Two possible moderators are the task requirements of the various AFSs and the characteristics of the individuals assigned to a particular AFS (e.g., predominantly male versus predominantly female incumbents). A third possible moderator variable is the form of the AFOQT used. The data from the current study could be re-analyzed by calculating validity coefficients for each form of the AFOQT separately. A meta-analysis of these newly calculated validity coefficients would allow for the determination of validity generalizability across the various AFOQT forms. If validity generalization was not supported across forms of the AFOQT, research could be conducted to determine what variables were moderating the relationship between AFOQT scores and technical school performance.

REFERENCES

Air Force Regulation 35-8. (1983, April). Air Force military personnel testing system (PA). Washington, DC: Department of the Air Force.

Arth, T. O. (1986, January). Validation of the AFOQT for non-rated officers (AFHRL-TP-85-50, AD-A164 134). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Finegold, L. S., & Rogers, D. (1985, June). Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training (AFHRL-TR-85-13, AD-A158 162). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Hunter, J. E., Schmidt, F. L., & Jackson, G. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage Publications.

Miller, R. (1960, September). Prediction of technical training criteria from AFOQT composites (WADD-TN-60-215, AD-246 658). Lackland AFB, TX: Personnel Laboratory, Wright Air Development Division.

Mosteller, F., & Bush, R. (1954). Selected quantitative techniques. In G. Lindzey (Ed.), Handbook of social psychology: Vol I. Theory and method. Cambridge, MA: Addison-Wesley.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. American Psychologist, 36, 1128-1137.